

---

## Corpus de référence et corpus d'usages : méthodologie de constitution pour une analyse des communications pilote-contrôleur

*Compiling a Reference Corpus and a Real Usage Corpus : Methodology for an  
Analysis of Pilot-Controller Communications*

**Stéphanie Lopez**

---



### Édition électronique

URL : <http://journals.openedition.org/praxematique/1112>

DOI : 10.4000/praxematique.1112

ISSN : 2111-5044

### Éditeur

Presses universitaires de la Méditerranée

### Édition imprimée

Date de publication : 1 janvier 2010

Pagination : 59-72

ISBN : 978-2-36781-012-6

ISSN : 0765-4944

### Référence électronique

Stéphanie Lopez, « Corpus de référence et corpus d'usages : méthodologie de constitution pour une analyse des communications pilote-contrôleur », *Cahiers de praxématique* [En ligne], 54-55 | 2010, mis en ligne le 01 janvier 2013, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/praxematique/1112> ; DOI : <https://doi.org/10.4000/praxematique.1112>

---

Tous droits réservés

Stéphanie Lopez

CLLE-ERSS U.M.R. 5263 (université Toulouse 2 — C.N.R.S.) et ENAC

---

## **Corpus de référence et corpus d'usages : méthodologie de constitution pour une analyse des communications pilote-contrôleur <sup>1</sup>**

### **Introduction**

Le domaine du contrôle aérien offre l'exemple parfait de l'instauration et de la mise en pratique d'une norme langagière prescriptive : la *phraséologie aéronautique*, un langage contrôlé créé pour couvrir les situations les plus courantes de la navigation aérienne afin de sécuriser et d'optimiser les communications radiotéléphoniques entre pilotes et contrôleurs. Lorsque ces derniers ne partagent pas la même langue maternelle, lors de situations de contrôle de vols internationaux par exemple, la phraséologie alors employée est basée sur la langue anglaise. Afin de garantir une communication radiotéléphonique concise, efficace et non-ambiguë, elle est soumise à des règles syntaxiques, lexicales, sémantiques et phonétiques strictes et ne laisse en principe aucune place à la créativité. Les situations de contrôle moins fréquentes et non couvertes par la phraséologie nécessitent, quant à elles, un recours à une forme langagière plus naturelle : le *plain language*.

Ces différentes formes langagières que sont la *phraséologie* et le *plain language*, ainsi que les variations qui en découlent, sont au cœur de notre projet de recherche en doctorat. Celui-ci a pour origine les besoins de l'École nationale de l'Aviation civile (ENAC) en matière d'enseignement de l'anglais. Notre objectif consiste à dresser un panorama des usages faits de la langue anglaise par les contrôleurs français

---

1. Cette étude est financée par l'École nationale de l'Aviation civile (ENAC) dans le cadre d'un projet de recherche en doctorat.

et les pilotes étrangers lors de leurs communications radiotéléphoniques. Pour cela, nous avons décidé de baser notre étude sur une analyse comparative entre un corpus de référence<sup>1</sup>, représentant la norme prescrite, et un corpus de communications pilote-contrôleur réelles, représentant les usages faits de celle-ci. Nous pensons en effet que les corpus sont essentiels à l'identification des éléments centraux et typiques d'une langue ou d'un langage (Sinclair, 1991) : ils ont pour objectif de « rendre compte de tous les aspects de la langue, à partir d'une collection qui en forme un échantillon raisonné » (Blanche-Benveniste, 2000). La méthodologie de constitution et de traitement d'un corpus joue donc un rôle primordial dans l'exploitation optimale des données recueillies.

Cet article vise à présenter la méthodologie mise en œuvre pour la constitution et le traitement de nos corpus afin de permettre, par la suite, une comparaison de nos données la plus juste possible. Nous introduirons dans un premier temps les notions de phraséologie et de *plain language*, ainsi que leurs caractéristiques. Nous présenterons ensuite les méthodes de constitution des corpus de référence et d'usages réels, avant de donner un bref aperçu du « nettoyage » appliqué aux corpus afin de favoriser une comparaison pertinente du lexique de ces derniers.

## 1. Phraséologie et *plain language*

Lors de situations de contrôle du trafic international, les communications entre pilotes d'aéronefs et contrôleurs de la navigation aérienne (ou communications sol-bord) s'effectuent principalement grâce à la *phraséologie aéronautique*. Il s'agit d'un langage contrôlé créé et modelé par l'Organisation de l'Aviation civile internationale<sup>2</sup>,

---

1. Il ne s'agit pas là d'un *corpus de référence* telle que cette notion est généralement définie et comprise par la communauté des linguistes. L'usage que nous faisons de *corpus de référence* n'exprime pas en premier lieu l'idée de complétude, mais insiste plutôt sur la dimension « obligatoire », « normative » de notre corpus : celui-ci vise à représenter la norme langagière prescrite dans un domaine spécialisé précis, autrement dit, les messages phraséologiques auxquels les pilotes et contrôleurs doivent se référer pour accomplir leurs activités de travail.

2. L'OACI (ou ICAO : *International Civil Aviation Organization*) est l'organisation dépendante des Nations Unies chargée de l'élaboration des normes à l'origine de la standardisation du transport aérien civil international.

d'après la langue anglaise, pour sécuriser de façon optimale les échanges radiotéléphoniques en simplifiant, clarifiant et désambiguïsant leur contenu. Pour cela, la phraséologie est composée de messages comprenant une structure syntaxique très particulière, un lexique et une prononciation spécifiques au domaine du contrôle aérien. Ses principales caractéristiques sont les suivantes (DGAC, 2007 ; Mell, 1992 ; Philps, 1989, 1991 ; Rubenbauer, 2009) :

- l'omniprésence de la tournure impérative dans les messages émis par les contrôleurs ;
- la rareté des tournures interrogatives et négatives ;
- la quasi-absence des auxiliaires modaux ;
- l'effacement systématique des déterminants (article défini ou pronom possessif), du pronom personnel sujet, des prépositions et des auxiliaires *be* et *have* dans les formes *be* + *-en*, *be* + *-ing* et *have* + *-en* ;
- la nominalisation des concepts ;
- l'emploi d'un lexique hautement spécialisé, univoque et fini (inférieur à 1 000 mots) ;
- l'emploi d'un alphabet propre à la radiotéléphonie (alpha, bravo, charlie, delta, echo, etc.) ;
- une épellation et une prononciation particulières des chiffres.

Certaines de ces caractéristiques sont illustrées dans les exemples, issus de notre corpus de référence, présentés ci-dessous :

*P<sup>1</sup> : Merignac Delivery, Citron Air 3 2 4 5, stand delta 8, requesting start-up, destination Lyon, information Lima.*

*C : Citron Air 3 2 4 5, start-up approved, CTOT 1 2 5 0, Sauveterre 3 alpha departure, level 1 1 0, squawk 2 3 2 0.*

*P : roger, start-up approved, CTOT 1 2 5 0, Sauveterre 3 alpha, level 1 1 0, squawk 2 3 2 0, Citron Air 3 2 4 5.*

*C : Citron Air 3 2 4 5, correct, report ready for push-back with Merignac Ground 1 2 1, 9.*

Ces caractéristiques syntaxiques, lexicales, sémantiques et phonétiques font de la phraséologie l'outil essentiel à la transmission des connaissances spécialisées nécessaires à une gestion sécurisée et optimale du trafic aérien. Cependant, la phraséologie est un outil limité : elle a été pensée pour couvrir un certain nombre de situations de contrôle. Par

---

1. Les énoncés de communications sol-bord commençant par « *P :* » sont émis par un pilote et ceux commençant par « *C :* » par un contrôleur.

conséquent, lors de situations non couvertes par celle-ci, les pilotes et les contrôleurs doivent avoir recours à une forme langagière plus naturelle, le *plain language*, défini par l'OACI (2010 : X<sup>1</sup>) comme « *the spontaneous, creative and non-coded use of a given natural language* ». Cette forme langagière, dont le recours lors de situations inhabituelles ou d'urgences est aisément admis, est tout de même utile à bon nombre de situations du quotidien (OACI, 2010 : 3.3.17). Ses principales caractéristiques sont les suivantes (OACI, 2010) :

- le *plain language* doit être employé dans les situations pour lesquelles la phraséologie n'est pas disponible, sans pour autant remplacer celle-ci complètement ;
- il comprend un lexique plus vaste que celui de la phraséologie pouvant être en rapport avec des domaines et des thèmes extérieurs à celui de l'aviation ;
- il doit être délivré, comme la phraséologie, de façon claire, précise et non-ambiguë afin de clarifier et d'élaborer des instructions ou lorsque le besoin de négocier des instructions se fait sentir ;
- il doit atteindre les mêmes objectifs que la phraséologie ;
- et ne doit donc, en aucun cas, être interprété comme un permis de bavarder sur la fréquence.

La définition et les caractéristiques du *plain language*, présentées ci-dessus, nous laissent penser que cette forme langagière s'apparente davantage au langage naturel qu'à un langage contrôlé. L'OACI (2010 : 6.2.8.4) le présente d'ailleurs comme un langage non-phraséologique : « *In simple terms, plain language can be thought of as the non-phraseology language that is used by participants in radiotelephony communications when standardized phraseology is not appropriate.* » Pourtant, le *plain language* semble être influencé par la phraséologie pour répondre aux besoins de concision, de clarté et de non-ambiguïté qui régissent les communications sol-bord (OACI, 2010 : 3.3.14) :

Plain language in aeronautical radiotelephony communications means the spontaneous, creative and non-coded use of a given natural language, although constrained by the functions and topics (aviation and non-aviation) that are required by aeronautical radiotelephony communications, as well as by specific safety-critical requirements for intelligibility, directness, appropriacy, non-ambiguity and concision.

---

1. Il s'agit bien ici d'une référence de pagination en chiffre romain.

Nous sommes donc en droit de nous demander si le *plain language* correspond au langage naturel employé dans un domaine restreint, celui du contrôle de la navigation aérienne — il s'agirait dans ce cas d'une langue de spécialité (Lerat, 1995) — ou s'il doit être considéré comme un langage contrôlé qui présenterait une structure plus naturelle que la phraséologie prescrite. Pour l'instant, il nous semble préférable de qualifier le plain language de forme langagière « plus ou moins » naturelle, ayant certaines des restrictions de la phraséologie et pouvant présenter certaines de ses caractéristiques.

Afin d'avoir une meilleure idée des formes langagières employées lors des communications sol-bord, et surtout de l'emploi du *plain language* par rapport à la phraséologie et au langage naturel, l'ENAC a initié, en collaboration avec le laboratoire CLLE-ERSS<sup>1</sup>, une observation des usages réels faits de la langue anglaise par les contrôleurs de la navigation aérienne français et les pilotes étrangers lors de leurs échanges radiotéléphoniques. Cette observation est réalisée par le biais d'un projet de recherche en doctorat basé sur une étude comparative entre deux corpus : un corpus de référence, représentant la norme prescrite, et un corpus de communications sol-bord réelles, représentant les usages. Ces deux corpus ont donc dû être définis et constitués d'après des critères spécifiques.

## 2. Constitution du corpus de référence

La première étape pour constituer notre corpus de référence a consisté en la sélection de textes desquels pouvaient être extraits des échantillons représentatifs du langage contrôlé qu'est la phraséologie aéronautique anglaise ou plus précisément, représentatifs de la norme prescrite (Bowker et Pearson, 2002). Ces textes devaient donc avoir une valeur de norme et devaient contenir des énoncés « complets » de phraséologie en anglais et non pas uniquement des listes de vocabulaires ou d'expressions conventionnelles. Peu de textes répondaient à ces critères, mis à part des manuels pédagogiques. Les exemples en

---

1. Cognition, langues, langage, ergonomie — Équipe de recherche en syntaxe et sémantique (U.M.R. 5263 — C.N.R.S. et université de Toulouse).

anglais issus de deux manuels<sup>1</sup> de formation à la phraséologie ont donc été sélectionnés pour constituer notre corpus de référence. Ces manuels sont le *Manual of Radiotelephony* (désormais *MoR*), édité par l'OACI (2007) et le *Manuel de formation à la phraséologie à l'usage de la circulation aérienne générale* (désormais *MFP*), édité par la direction générale de l'Aviation civile<sup>2</sup> (DGAC, 2007). Nous espérons ainsi représenter respectivement la norme d'un point de vue international — celle à laquelle certains pilotes étrangers peuvent se référer — et la norme d'un point de vue national — à laquelle tous les contrôleurs de notre corpus d'usages réels sont soumis<sup>3</sup>.

Notre choix de constituer un corpus de référence avec des exemples issus de manuels pédagogiques n'est pas commun. Les raisons qui nous ont confortée dans notre décision sont les suivantes :

Tout d'abord, un manuel est un « ouvrage [...] qui contient les éléments d'une discipline ou l'essentiel d'un programme scolaire » (TLFI). Un manuel devrait donc pouvoir être considéré comme exhaustif : les manuels choisis tentent, en effet, de couvrir la majorité des aspects phraséologiques des communications radiotéléphoniques. Ainsi, les 1 395 exemples de « prises de parole » en anglais que le *MoR* et le *MFP* regroupent à travers leurs différents chapitres peuvent être considérés comme représentatifs d'une phraséologie internationale fondamentale.

De plus, les manuels choisis sont tous deux des références en matière de présentation et d'enseignement de la phraséologie : le *MFP*, qui s'inspire essentiellement du *MoR* (DGAC, 2007), est considéré en France comme le « document de référence pour la formation initiale et continue du personnel des organismes de la circulation aérienne à l'École nationale de l'Aviation civile (ENAC), dans les centres en-route, les approches et les aérodromes » (DGAC, 2007 : 3). Les deux

---

1. Les manuels étant constitués, entres autres, de titres, de consignes, et pour le manuel français, d'exemples en français comme en anglais, seuls les exemples d'énoncés en anglais ont été sélectionnés.

2. La DGAC est l'administration française en charge de la réglementation, de la supervision, de la sécurité et de la sûreté du transport aérien en France.

3. La phraséologie peut être modifiée à l'échelle nationale, sous réserve d'en informer officiellement l'OACI. En France, où elle a une valeur officielle de norme, très peu de modifications ont été apportées à la phraséologie recommandée par l'OACI. Ce n'est pas forcément le cas de tous les pays. Les pilotes étrangers peuvent donc avoir pour référence une phraséologie comprenant des différences avec celle appliquée en France.

manuels peuvent donc être considérés comme reflétant une dimension normative.

Tous deux sont bien sûr en conformité avec les procédures de radiotéléphonie internationales (et nationales en ce qui concerne le *MFP*) en vigueur et proposent une illustration de celles-ci par des exemples : « *The purpose of this manual is to provide examples of the radiotelephony phraseology [...]* » (OACI, 2007 : iii) ; « il a pour objectif de proposer, par des exemples, au personnel des organismes de la circulation aérienne une phraséologie de référence correspondant aux situations les plus courantes [...] » (DGAC, 2007 : 3). Ces exemples, qui reflètent une dimension normative du fait du caractère officiel et prescriptif des textes dont ils sont issus, peuvent ainsi être considérés comme un modèle à suivre, à imiter : ce sont des représentations orthographiques des usages qu'il faudrait employer sur la fréquence radiotéléphonique.

D'ailleurs, les définitions générale et spécialisée de l'exemple « action, manière d'être, considérée comme pouvant être imitée » (Rey-Debove & Rey, 2009) et « passage d'un texte, phrase ou membre de phrase que l'on cite à l'appui d'une explication pour illustrer l'emploi d'un fait linguistique » (Rey-Debove & Rey) corrobore notre représentation de l'exemple et nous incite à traiter ceux constituant notre corpus de référence comme une illustration des règles de la phraséologie prescrite correctement appliquées.

Les différentes raisons évoquées ci-dessus nous permettent de considérer notre corpus de référence comme représentatif de la phraséologie anglaise prescrite. Le fait que ce corpus soit constitué d'exemples écrits alors que la phraséologie est, d'ordinaire, un langage verbal ne nous semble pas problématique car les exemples issus des manuels sont censés être la représentation orthographique d'énoncés oraux. De plus, la comparaison souhaitée entre le corpus de référence et celui d'usages réels concerne la transcription orthographique de ce dernier (voir plus bas).

### 3. Constitution du corpus d'usages réels

Le corpus d'usages réels, comme son nom l'indique, devait avant tout être représentatif des usages langagiers produits sur la fréquence radiotéléphonique lors de situations de contrôle réelles. Il devait



également être représentatif des trois types de contrôle existants<sup>1</sup>, représentés dans les manuels, et être assez récent et volumineux. Celui-ci devait donc être constitué d'enregistrements de communications réelles, être réalisé auprès de différents centres de contrôle et concerner plusieurs secteurs et créneaux horaires.

#### 4. Collecte des données sonores

L'obtention d'enregistrements de communications sol-bord réelles a nécessité une autorisation officielle au préalable, car en France ce genre de données n'est pas accessible au grand public. Une fois l'autorisation de la DGAC obtenue, plusieurs sous-corpus ont pu être constitués par nos soins, en collaboration avec trois centres de contrôle de la navigation aérienne français : deux CRNA et un grand aéroport, choisis principalement pour leur taux d'occupation de la fréquence en anglais ainsi que pour leur intérêt pour notre projet de recherche. Dans un souci d'anonymat, nous ne précisons pas ici de quels centres il s'agit exactement (une des conditions *sine qua non* à l'obtention des données sonores était en effet l'anonymisation des locuteurs concernés par les enregistrements et des centres dont ils dépendent). Plus de 70 heures d'enregistrement de communications sol-bord, réalisées lors de situations de contrôle routinières dans des secteurs différents et à des créneaux horaires variés, ont donc été récupérées. Ces enregistrements concernent des activités de contrôle dans leur contexte ordinaire, c'est-à-dire « telles qu'elles ont lieu de manière ordinaire, même en l'absence du chercheur » (Baude, 2006 : 50). Il s'agit plus précisément de « reprise[s] d'enregistrement » (*ibid.*) car en France, les communications sol-bord sont enregistrées en permanence et stockées pendant plusieurs semaines par les centres de contrôle eux-mêmes afin de pouvoir procéder, si besoin, à une analyse de celles-ci *a posteriori*. Nous n'avons donc pas eu à choisir de dispositif d'enregistrement : bénéficier des dispositifs intégrés aux postes de contrôle nous a permis de ne pas manipuler d'appareil en présence des contrôleurs et ainsi de minimiser l'influence fâcheuse que ce genre d'instrument peut

---

1. Les trois types de contrôle correspondent aux différentes phases de vol d'un aéronef. En France, le *contrôle d'Aérodrome* et le *contrôle d'Approche* sont pris en charge par les aéroports tandis que le *contrôle en Route* est effectué par les Centres en route de la navigation aérienne (CRNA).

avoir sur les productions verbales des enquêtés (Bowker et Pearson, 2002 ; Baude, 2006). De plus, les contrôleurs se sachant inéluctablement enregistrés lorsqu'ils sont en poste, nous espérons avoir obtenu des données présentant un certain degré de spontanéité.

Afin de permettre une comparaison outillée des données des deux corpus, la transcription orthographique d'un certain nombre d'enregistrements collectés était nécessaire.

## 5. Transcription orthographique

Un protocole de transcription orthographique <sup>1</sup> a été défini et appliqué aux seules communications en anglais contenues dans un peu plus de 22 heures d'enregistrements <sup>2</sup>. Pour cela, nous avons opté pour l'outil d'aide à la transcription PRAAT ([www.fon.hum.uva.nl/praat/](http://www.fon.hum.uva.nl/praat/) et Delais-Roussarie, Meqqori et Tarrier, 2003). Les heures de communications à transcrire ont été sélectionnées afin de représenter différentes situations de contrôle : différents types de contrôle, différents secteurs, différents locuteurs et différents créneaux horaires (comprenant des périodes de trafic chargées, des périodes creuses, des créneaux de jour et de nuit).

Bien sûr, la transcription de communications sol-bord, qui plus est en anglais, nécessite une connaissance au préalable de la phraséologie et du domaine du contrôle aérien même pour des locuteurs anglophones natifs. Malgré ces connaissances, la tâche de transcription reste « éminemment interprétative, en particulier au niveau orthographique » (Véronis, 2000 : 97). C'est pourquoi toutes les transcriptions ont été soumises à une correction et une validation de la part de contrôleurs aériens, experts du domaine. Ces derniers ne sont bien sûr pas à l'abri d'une mauvaise interprétation des énoncés enregistrés : nous visons principalement, par cette double vérification, à ce que la transcription réalisée ne reflète pas nos interprétations.

---

1. La plupart des conventions de transcription choisies sont inspirées des conventions régissant les corpus de la banque de données *Valibel* et des réflexions qui en ont découlées (DISTER *et al.*, 2004 ; DISTER et SIMON, 2008).

2. Une heure d'enregistrement n'équivaut pas à une heure de communication : les échanges radiotéléphonique entre pilotes et contrôleurs sont en effet généralement séparés de pauses ou silences qui peuvent durer entre quelques secondes et plusieurs minutes.

Notre méthodologie de constitution du corpus d'usages, qu'il s'agisse de la collecte des données sonores ou de la transcription orthographique de celles-ci, vise une représentativité des communications sol-bord en anglais observables sur la fréquence radiotéléphonique. Nous espérons ainsi que les analyses conduites sur nos corpus permettront de dresser un panorama représentatif des emplois de la phraséologie et du *plain language*, ainsi que des variations qui en découlent. Afin de procéder, dans un premier temps, à une analyse comparative pertinente du lexique des corpus, nous avons effectué un nettoyage des données de ces derniers.

## 6. Nettoyage des données

Commençons par préciser ici que la structure syntaxique des énoncés qui composent nos corpus étant très particulière, aucune annotation automatique suffisamment correcte n'a pu être réalisée. La phase de correction manuelle des erreurs d'étiquetage, généralement nécessaire mais relativement légère (Véronis, 2000), aurait été trop conséquente si nous avions procédé à un étiquetage morpho-syntaxique outillé des données. Aussi, nos corpus sont des « corpus nus » (Blanche-Benveniste, 2000) et les observations sont effectuées à l'aide du concordancier *Antconc* (Anthony, 2004), qui nous permet également de connaître les dimensions de nos corpus : celui de référence est constitué de 11 844 occurrences comprenant 805 types de mots différents et celui d'usages comprend 49 020 occurrences et 1 238 types. Le faible nombre de types de mots contenus dans les corpus s'explique par le fait que le vocabulaire de la phraséologie, comme tout langage opératif, est restreint par rapport au lexique de la langue naturelle (Falzon, 1986). Il s'agit là d'une des caractéristiques principales de la phraséologie introduites plus haut.

Dans le but de recenser et de comparer le lexique des deux corpus, un « nettoyage » de ces derniers était de rigueur : cette démarche nous permet de ne prendre en compte que le lexique fondamental au contrôle aérien. Les formes lexicales écartées de nos analyses, ainsi que les raisons de leur exclusion, sont présentées ci-dessous :

- les disfluences (amorces de mots, pauses pleines, etc. : *corral*, *behinl*, *col*, *-huh-*), caractéristiques des communications verbales,

- présentes dans le corpus d'usages réels — le corpus de référence ne contenant aucune disfluen­ce ;
- les marqueurs de politesse ou de salutation énoncés dans d'autres langues que l'anglais (*bonjour, Tchüss, gracias, konichiwa*, etc.) éparpillés dans le corpus d'usages réels au gré de la créativité langagière des pilotes et des contrôleurs – le corpus de référence n'étant constitué que d'énoncés en anglais ;
  - les lettres de l'alphabet (*alpha, bravo, charlie, delta*, etc.) contenues dans les deux corpus — les 26 lettres n'étant pas toutes recensées dans le corpus de référence ;
  - les noms propres (noms authentiques ou fictifs de villes, d'aéroports, d'aéronefs, de balises d'aide à la navigation, etc.) présents en grande quantité dans les deux corpus — ces derniers ne pouvant pas être considérés comme spécifiques au contrôle aérien ;
  - les hapax (les mots n'ayant qu'une seule occurrence) présents dans le corpus d'usages réels<sup>1</sup> (*airport, avoided, enable, faster, luggage, whichever, whoops*, etc.) — ceux-ci ne pouvant pas être considérés comme représentatifs des usages ;

Après nettoyage, le corpus de référence ne comprend plus que 671 types de mots (au lieu de 805) et celui d'usages 492 (au lieu de 1 238). Une comparaison préliminaire nous permet de savoir que les deux corpus ont 284 types de mots en commun et que le corpus de référence et le corpus d'usages ont respectivement 387 et 208 types de mots qui leur sont spécifiques. Ces différents types de mots constitueront la base de notre analyse comparative du lexique des deux corpus. On pourra par exemple se demander à quelles catégories grammaticales correspondent les 208 types spécifiques au corpus d'usages et si certaines catégories sont sous-représentées ou au contraire surreprésentées par rapport aux catégories grammaticales présentes dans le corpus de référence. Selon nous, le nettoyage des corpus nous permet de focaliser nos analyses sur les formes lexicales essentielles au contrôle de la navigation aérienne, que celles-ci correspondent à la norme prescrite ou aux usages qui en sont faits. Seule une observation, une catégorisation et une comparaison davantage détaillées des différents types de mots contenus dans les corpus nettoyés nous permettra de dresser

---

1. Les exemples, ayant un statut particulier de « modèle », nous avons choisi de conserver les hapax du corpus de référence.

un panorama représentatif des formes langagières employées sur la fréquence radiotéléphonique.

## Conclusion

Dans cet article, nous avons présenté, à travers notre expérience, l'importance de la méthodologie de constitution de corpus, que les données soient écrites ou orales, d'après des critères spécifiques et explicites : la représentativité et l'exhaustivité du corpus en dépendent.

Dans notre cas, le choix précis des textes desquels sont extraits les exemples composant le corpus de référence et la méthodologie de récupération des enregistrements oraux et de transcription orthographique constituant le corpus d'usages réels, nous permettent d'obtenir deux corpus représentatifs respectivement de la norme langagière prescrite dans le domaine du contrôle aérien et des usages réels qui sont faits de celle-ci. Le nettoyage appliqué aux données permet, dans un premier temps, une analyse comparative des corpus d'un point de vue lexical ; mais par la suite, une analyse comparative à d'autres niveaux linguistiques — syntaxique, sémantique et phonétique — nous permettra d'avoir une meilleure idée des emplois de la phraséologie et du *plain language* dans les communications sol-bord en anglais.

## Références bibliographiques

- ANTHONY L., 2004, « Antconc : a learner and classroom friendly, multi-platform corpus analysis toolkit », in ANTHONY L., FUJITA S. & HARADA Y. (dir.), *Proceedings of IWLeL 2004 : an interactive workshop on language e-learning*, 7-13.
- BAUDE O. (dir.), 2006, *Corpus Oraux. Guide des bonnes pratiques*, Paris, C.N.R.S. éditions & Presses universitaires d'Orléans.
- BLANCHE-BENVENISTE C.,  
2000, « Introduction », in BILGER M. (dir.), *Corpus. Méthodologie et applications linguistiques*, Paris, Champion, 11-15.
- BOWKER L. & PEARSON J.,  
2002, *Working with specialized language, a practical guide to using corpora*, Londres, Routledge.
- DELAIS-ROUSSARIE E., MEQQORI A. & TARRIER J.-M.,  
2003, « Annoter et segmenter des données de parole sous PRAAT », in DELAIS-ROUSSARIE E. & DURAND J. (dir.),

*Corpus et variation en phonologie du français : méthodes et analyses*, Toulouse, Presses universitaires du Mirail, 159-185.

- DIRECTION GÉNÉRALE DE L'AVIATION CIVILE (DGAC),  
2007, *Phraséologie. Manuel de formation à la phraséologie à l'usage de la circulation aérienne générale*, Mérignac, Service de l'information aéronautique (SIA), récupéré le 10 septembre 2008 du site de l'éditeur : [www.sia.aviation-civile.gouv.fr/asp/frameset\\_fr.asp?m=26](http://www.sia.aviation-civile.gouv.fr/asp/frameset_fr.asp?m=26).
- DISTER A., FRANCARD M., GERON G., GIROUL V., HAMBYE P.,  
SIMON A.-C. & WILMET R., 2004, *Conventions de transcription régissant les corpus de la banque de données VALIBEL*, récupéré le 14 avril 2009 du site du centre de recherche VALIBEL : [www.uclouvain.be/cps/ucl/doc/valibel/documents/conventions\\_valibel\\_2004.PDF](http://www.uclouvain.be/cps/ucl/doc/valibel/documents/conventions_valibel_2004.PDF).
- DISTER A. & SIMON A.-C.,  
2008, « La transcription synchronisée des corpus oraux. Un aller-retour entre théorie, méthodologie et traitement informatisé », *Arena Romanistica* 1, 1, 54-79.
- REY-DEBOVE J. & REY A. (dir.),  
2009, *Le nouveau petit Robert de la langue française*, Paris, Le Robert.
- FALZON P.,  
1986, *Langages opératifs et compréhension opérative*, thèse de doctorat non publiée, université de Paris V — la Sorbonne.
- LERAT P.,  
1995, *Les langues spécialisées*, Paris, Presses universitaires de France.
- MANUEL,  
(n.d.), in *Le trésor de la langue française informatisé (TLFI)*, récupéré le 20 mars 2011 de <http://atilf.atilf.fr/>.
- MELL J.,  
1992, *Étude des communications verbales entre pilote et contrôleur en situation standard et non-standard* (thèse de doctorat non publiée), université de Toulouse II — le Mirail.
- ORGANISATION DE L'AVIATION CIVILE INTERNATIONALE (OACI),  
2007, *Manual of Radiotelephony. Doc 9432-AN/925* (4<sup>e</sup> éd.), Montréal, Canada.
- ORGANISATION DE L'AVIATION CIVILE INTERNATIONALE (OACI),  
2010, *Manual on the implementation of ICAO language proficiency requirements. Doc 9835-AN/453* (2<sup>e</sup> éd.), Montréal, Canada.

- PHILPS D., 1989, *L'anglais de la circulation aérienne*, Toulouse, ENAC.
- PHILPS D., 1991, « Linguistic security in the syntactic structures of air traffic control English », *English World-Wide*, 12, 1, 103-124.
- RUBENBAUER F., 2009, *Linguistics and flight safety : Aspects of oral English communication in aviation*, Aachen, Shaker Verlag.
- SINCLAIR J. M., 1991, *Corpus, concordance, collocation*, Oxford, Oxford University Press.
- VÉRONIS J., 2000, « Annotation automatique de corpus : état de la technique », in BILGER M. (dir.), *Corpus. Méthodologie et applications linguistiques*, Paris, Champion, 94-106.